

TRAVAUX DE L'INSTITUT DE PHONETIQUE D'AIX

Vol. 9, 1983/84 - Pp. 357-372

DESCRIPTION D'UN ANALYSEUR DE SONIE
DE LA PAROLE CONTINUE *

BERNARD TESTON

INSTITUT DE PHONETIQUE, L.A. 261

* Communication présentée au 11e Congrès International d'Acoustique, Paris, juillet 1983.

Description d'un analyseur de sonie de la parole continue

Résumé

Cet analyseur permet d'obtenir le tracé de l'évolution en fonction du temps de l'intensité sonore subjective sur des signaux de parole. Les pondérations effectuées sur la pression acoustique objective prennent en compte l'isosonie, l'effet de masque fréquentiel, la répartition spectrale de l'énergie ainsi que la durée des segments vocaliques et les effets de masque temporels. L'appareil est constitué par un analyseur de fréquence en bandes critiques et de circuits particuliers qui simulent ces diverses pondérations en temps réel. Sa dynamique est de 60 dB. Notre objectif est de faire ressortir la potentialité en tant qu'indice du paramètre d'intensité en vue d'une interprétation linguistique.

Description of a loudness analyzer for continuous speech

Abstract

This analyzer provides a plotting of the evaluation of the subjective intensity level of speech signals as a function of time. Weighting of the objective acoustic pressure level takes into account equal loudness, frequency masking, the distribution of energy in the spectrum as well as the duration of vocalic segments and temporal masking. The apparatus is composed of a critical band frequency analyzer and special circuits simulating the various weightings in real time. The range of measurement is 60 dB. Our aim is to bring to light the nature of the intensity parameter as a potential cue for linguistic interpretation.

DESCRIPTION D'UN ANALYSEUR DE SONIE DE LA PAROLE CONTINUE

Bernard TESTON

Introduction

Le paramètre acoustique de l'Intensité du signal de parole a toujours été pris en compte par les phonéticiens comme indice de lieu et de mode des unités phoniques. Paradoxalement, il a été bien moins étudié du point de vue prosodique si ce n'est dans le cas des accents expressifs, où son importance est généralement considérée comme secondaire en comparaison avec la mélodie et la durée.

On peut considérer les travaux de Rossi (1971-1978) comme à la charnière de ces deux points de vue, et ceux de Sorin (1981) comme une importante étape du point de vue prosodique.

Il est possible de réaliser des études prosodiques sur des données objectives mais cette procédure ne peut pas rendre compte des phénomènes linguistiques. En effet, seules les caractéristiques acoustiques des éléments phonétiques effectivement perçus peuvent avoir une valeur linguistique. Il est donc nécessaire d'effectuer un transcodage perceptif de ces caractéristiques en se

fondant sur des résultats psychoacoustiques et psycholinguistiques.

L'intensité effectivement perçue, ou sonie, intervient au niveau de l'intonation dans le cadre de l'énoncé dont l'étendue peut aller du groupe intonatif à la phrase et même au-delà de la phrase. Elle intervient également au niveau de l'accent dont l'étendue peut aller de la syllabe au mot et au groupe accentuel. La phrase n'est pas une unité de perception. C'est pour cette raison que les travaux sur la sonie globale de la phrase comme ceux de Fastl (1976) (2) ne peuvent satisfaire le phonéticien. Une approche analytique traitant la sonie de la parole continue au niveau des unités phoniques est nécessaire. Pour cela, nous cherchons à appliquer des procédures automatiques de correction de l'intensité en fonction des mécanismes de l'audition afin d'évaluer sa potentialité en tant qu'indice, en vue d'une interprétation linguistique. C'est le but que nous poursuivons partiellement avec l'analyseur de sonie que nous allons décrire, à la lumière de résultats psycho-acoustiques bien fondés.

Nous entendons par analyseur de sonie, un sonomètre doté de fonctions de pondération subjective, simulant certains mécanismes de l'audition et capable de ce fait d'effectuer une mesure objective du niveau d'excitation à partir du niveau de pression sonore. Nous l'avons réalisé dans le but de mesurer le niveau d'excitation du signal de parole au cours de son développement, donc en temps réel. La sonie d'un stimulus auditif est directement fonction de la valeur efficace de la pression acoustique au tympan. Pour définir le niveau d'excitation à partir de cette valeur, il est nécessaire de prendre en compte les influences des mécanismes auditifs, d'une part dans le domaine fréquentiel, tels que l'isosonie, l'effet de masque fréquentiel, l'effet de la répartition de l'énergie dans le spectre, et, d'autre part dans le domaine temporel, tels que la durée des segments et les effets de masque temporels.

Les pondérations fréquentielles

De nombreuses méthodes d'évaluation du niveau d'excitation à partir de la connaissance du niveau de pression et du spectre du stimulus ont été proposés. Certaines sont normalisées par l'ISO depuis 1966 (Zwicker 1960 et Stevens 1961) pour des bruits continus ou se reproduisant à intervalles réguliers. Ce n'est que plus récemment que l'on a proposé des calculs tenant compte des variations temporelles des signaux (Zwicker 1977).

Dans une étude sur l'intensité spécifique des voyelles du français, Rossi en 1971 montre que ce phénomène peut être expliqué par une interprétation psychoacoustique du signal. Il en déduit une méthode objective de calcul dérivée des méthodes normalisées. Après l'application des monogrammes de Zwicker (1960), (qui prennent en considération une analyse par bandes critiques en fonction de l'isosonie et de l'effet de masque fréquentiel), Rossi (1971) applique une correction en fonction de l'écartement des deux premiers formants sur le spectre, au moyen d'un facteur F , dérivé de la méthode de Stevens (1961). C'est cette procédure que nous utilisons pour le calcul automatique des pondérations fréquentielles (Figure 1).

Bien que la perception d'objets sonores à spectres harmoniques tels que les voyelles soit identique à celle d'objets sonores à spectres continus tels que les consonnes constrictives (Espesser et al. 1982), la méthode de Rossi (1971) s'applique assez mal au spectre large de ces dernières. Cependant, pour des raisons de simplicité, nous l'utilisons sur toute la chaîne parlée sans distinction de catégorie des éléments phoniques.

La pondération temporelle

Le signal de parole est fortement modulé en fonction du temps. Ses variations d'énergie proviennent des différences in-

trinsèques entre les segments contigus, des effets du contexte et des facteurs phonotactiques tels que le nombre de syllabes composant le mot, le syntagme ou la phrase et prosodiques tels que l'accent, l'intonation, le rythme, la pause et le débit. Pour ces raisons, la sonie de la parole continue est fortement marquée par l'évolution temporelle du signal, essentiellement en fonction de la durée des segments et des effets de masque temporels. Pour tenir compte de ces phénomènes nous nous inspirons des travaux de Zwicker de 1977 pour la simulation de l'effet de masque temporel et des résultats de Rossi (1978) et Sorin (1981) sur l'intégration temporelle des variations d'intensité.

L'analyseur de sonie

Il est essentiellement constitué par un analyseur de fréquence en temps réel dont la structure originale remonte aux travaux de Pfeiffer (1965). Il dispose de 25 canaux d'analyse qui simulent les bandes critiques dans la bande passante de l'oreille. Chaque filtre d'analyse est constitué par un filtre passe haut de 60 dB/octave associé à un filtre passe bas de 24 dB/octave afin de tenir compte de l'effet de masque fréquentiel. Les filtres sont de Butterworth pour optimiser leur temps de propagation de groupe. Le niveau d'énergie de chaque bande est pondéré en fonction de l'isosonie. Il est détecté au moyen d'un convertisseur de valeur efficace d'un facteur de crête de 3 et dont la constante de temps d'intégration exponentielle moyenne pondérée est de 10 ms pour les deux bandes les plus basses et de 5 ms pour les autres. Les canaux d'analyse sont réalisés en techniques analogiques. Après multiplexage, la conversion logarithmique et le calcul du facteur F sont réalisés en techniques numériques. Parallèlement à l'analyseur, nous avons un détecteur de l'énergie globale du signal de parole identique à celui des canaux d'analyse. Il n'en diffère que par un facteur de crête plus élevé (8), sa constante de temps est de 10 ms. La valeur de l'énergie globale est modulée par le signal résultant des diverses pondérations fréquentielles (Figure 2).

La pondération en fonction de l'effet de masque temporel est effectuée au moyen de fronts décroissants exponentiels proportionnels à la durée du signal selon une méthode décrite par Vogel en 1975 et en accord avec les résultats de Fastl (1976 b) sur l'effet de masque temporel en fonction de la durée de bruits à large bande (Figure 3 a).

La pondération temporelle en fonction de la durée du signal tient compte des résultats de Zwislocki en 1969 et Pedersen et al. en 1977. Elle correspond à une durée critique de 200 ms qui est également la valeur du temps d'intégration en fin de phrase décelée par Sorin en 1981. Cette durée correspond à une constante de temps exponentielle pondérée (RC) de 80 ms (Figure 3 b, et c).

Les courbes d'intensité en dB et de sonie en phones ainsi obtenues sont exploitées au moyen d'un enregistreur galvanométrique. La dynamique de l'analyseur est de 60 dB ou Phones. Les spectres sont échantillonnés toutes les 10 ms (Figures 4 et 5).

CONCLUSION

Nous envisageons d'améliorer ces résultats dans deux directions. D'une part, au moyen d'études particulières sur des stimulus de parole tels que la syllabe (Nishinuma et al. 1983), ou également au niveau de la phrase avec des procédures de tests appropriées. Nous nous employons dans ce sens. D'autre part en affinant la pondération fréquentielle ce qui peut être réalisé suivant deux procédures : la méthode de Rossi (1971) pour les éléments phoniques sonores et une méthode qui nous reste à définir pour les consonnes non voisées. Le choix des méthodes étant effectué au moyen d'un détecteur de voisement en temps réel.

BIBLIOGRAPHIE

- ESPESSER, R. & TESTON, B. (1982), "Loudness : Continuous spectrum sound versus harmonic spectrum sound", **Forschritte des Akustic - FASE/DAGA '82**, 1091-1094.
- FASTL, H. (1976 a), "Temporal masking effects : I. Broad band noise masker", **Acustica**, 35, 287-302.
- FASTL, H. (1976 b), "Schallpegel und Lautstärke von Sprache", **Acustica**, 35, 341-345.
- ISO R532 (1966) : "Method for calculating loudness level".
- NISHINUMA, Y., DI CRISTO, A. & ESPESSER, R. (1983), "Loudness as a function of vowel duration in CV syllables", **Speech Communication**, 2, 167-169.
- PEDERSEN, O. J., LYREGAARD, P. E. & POULSEN, T. (1977), "The Round Robin test on evaluation of loudness level of impulsive noise", **Report No 22**, Acoustic Laboratory, Technical University of Denmark, ISO/TC43/SC1, 180 p. + Annex.
- PFEIFFER, T. (1966), "Ein Lautstärke-Messgerät für breitbandige und impulshaltige Schalle", **Acustica**, 17, 322-334.
- ROSSI, M. (1971), "L'intensité spécifique des voyelles", **Phonetica**, 24, 129-161.
- ROSSI, M. (1978), "The perception of non-repetitive intensity glides on vowels", **Journal of Phonetics**, 6, 384-396.
- SORIN, C. (1981), "Functions, roles, and treatments of intensity in speech", **Journal of Phonetics**, 9, 359-374.

STEVENS, S. S. (1961), "Procedure for calculating loudness : MARK 6", **J.A.S.A.**, 33, 1577-1585.

VOGEL, A. (1975), "Ein gemeinsames Funktionsschema zur Beschreibung der Lautheit und der Rauigkeit", **Biol. Cybernetics**, 18, 31-40.

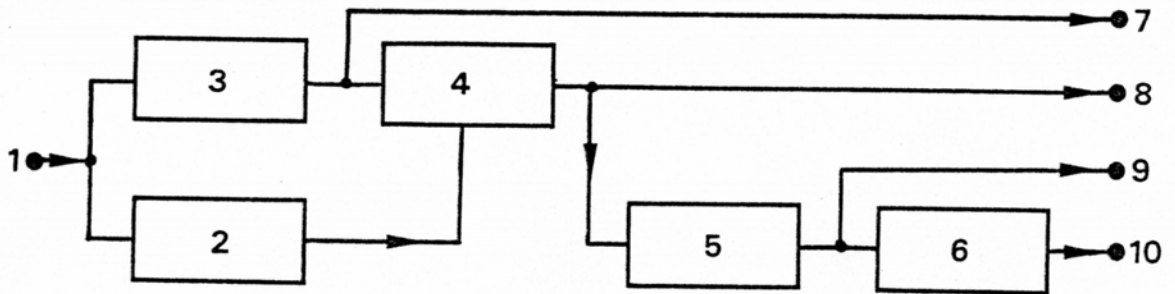
ZWICKER, E. (1960), "Ein Verfahren zur Bestimmung der Lautstärke", **Acustica**, 10, 304-310.

ZWICKER, E. (1977), "Procedure for calculating loudness of temporally variable sounds", **J.A.S.A.**, 62, 675-682.

ZWISLOCKI, J. J. (1969), "Temporal summation of loudness : An analysis", **J.A.S.A.**, 46, 431-441.

Voyelles	a	ã	ɔ	œ	o	ε	u	y	e	i
Isp observée	0	1,25	1	1,25	1,7	1,75	2,35	2,5	4	4,5
Phonie	72,2	73,2	72,8	73	73,9	73,9	74,7	75,2	76,4	77,3
Isp calc. (ROSSI)	0	1	0,6	0,8	1,7	1,7	2,5	3	4,2	5,1
Isp mes. (Analys.)	0	2	1,5	1,5	2	2,5	2,5	3	3,5	5,5

Figure 1 : Tableau comparatif des intensités spécifiques calculées par la méthode de Rossi (1971) et mesurées au moyen de l'analyseur de sonie pour différentes voyelles.



1. Entrée du sonomètre (signal de la pression acoustique). Ligne 1 sur les figures 4 et 5. - 2. Analyseur de fréquence en bandes critiques. - 3. Détecteur de la valeur efficace de la pression (constante de temps de 10 millisecondes). Conversions logarithmique. - 4. Modulateur de pondération en fonction des caractéristiques fréquentielles du signal. - 5. Pondération temporelle de décroissance de l'énergie du signal (fonction de la durée des segments). - 6. Pondération temporelle de croissance de l'énergie du signal (80-ms). - 7. Sortie du logarithme de la valeur efficace de la pression acoustique en dB. Ligne 2 des figures 4 et 5. - 8. Sortie du logarithme de la valeur efficace de la pression acoustique pondérée en fonction des caractéristiques fréquentielles du signal en Phones. Ligne 3 des figures 4 et 5. - 9. Même sortie qu'en 8 avec une pondération temporelle en fonction de la durée des segments (en Phones). Ligne 4 sur la figure 5. - 10. Même sortie qu'en 9 avec une pondération temporelle fonction de la croissance de l'énergie du signal (en Phones). Ligne 5 sur la figure 5.

Figure 2 : Schéma de principe de l'analyseur de sonie.

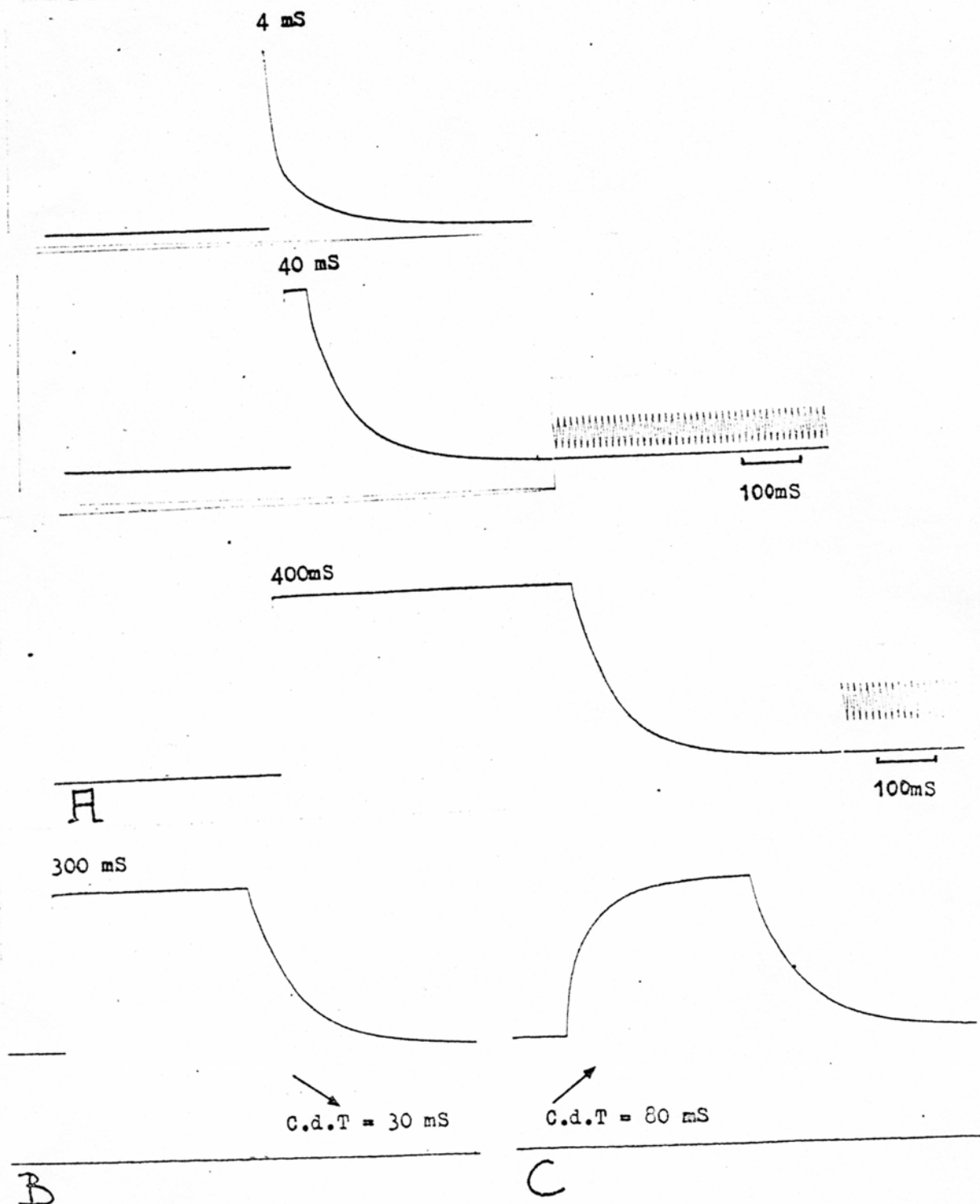


Figure 3 :

- a) Constante de temps de décroissance variable en fonction de la durée du signal pour 4, 40 et 400 ms.
- b) Constante de temps de décroissance pour une durée de 300 ms du stimulus. La valeur est dans ce cas de 30 ms (RC).
- c) Constante de temps de croissance. Sa valeur est de 80 ms. La constante de temps de décroissance est toujours de 30 ms pour une durée de stimulus de 300 ms.

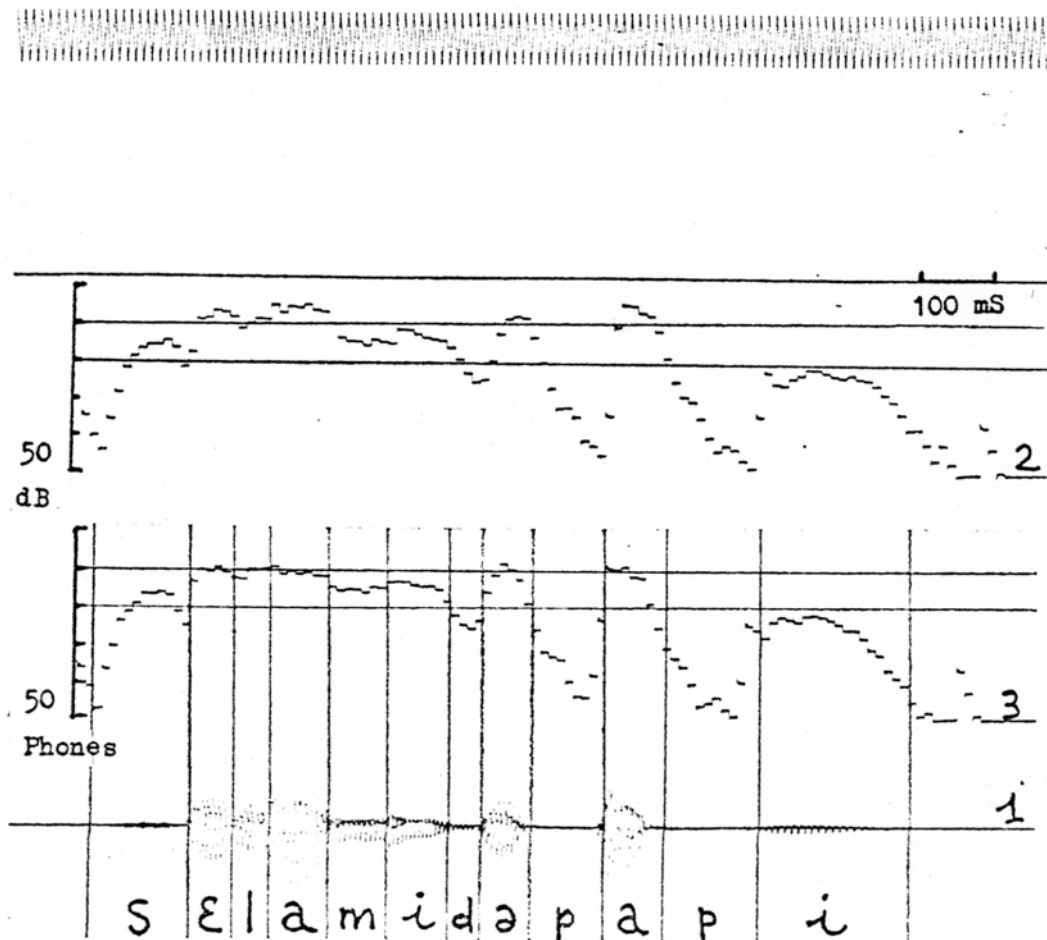


Figure 4 : Exemple de pondération automatique en fonction des paramètres fréquentiels au moyen de la méthode de Rossi (1971) sur la phrase : "C'est l'ami de Papi".

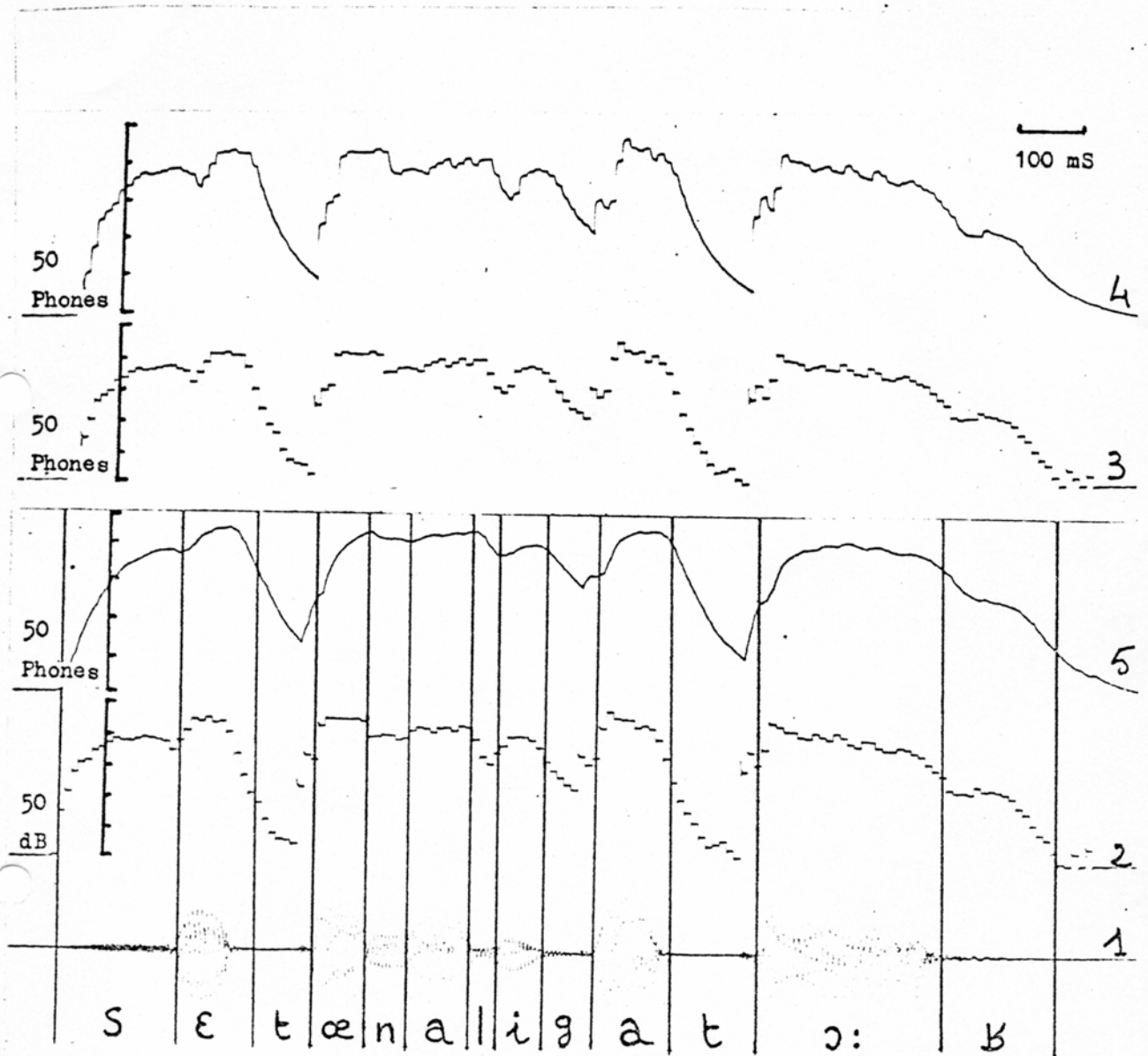


Figure 5 : Exemple de pondération automatique
en fonction des paramètres fréquentiels
et temporel sur la phrase :
"C'est un alligator".